CLMPTO

01/30/04

MBL

CLAIMS 1-41 CANCELED

CLAIMS 42-72 ADDED

Claims 1-41 (canceled)

Claim 42 (new): A method of partitioning data records in a computer into groups, comprising the steps of:

(a) defining a function of a distribution of the values of a designated variable associated with the data records, wherein the function comprises a combination of measures;

(b) partitioning the values of a designated variable into two or more groups, wherein the value of the function is determined by applying an optimization procedure; and

(c) assigning each data record to a group according to the value of the designated variable.

Claim 43 (new): A method as recited in claim 42 wherein said partitioning comprises partitioning of data records into groups of approximately equal size.

Claim 44 (new): A method as recited in claim 42 further comprising the step of selecting a

partition from many computed solutions yielding acceptable performance.

Claim 45 (new): A method as recited in claim 42 wherein said optimization procedure results in an optimal assignment.

Claim 46 (new): A method as recited in claim 42 wherein said function is a combination of entropy and adjacency.

Claim 47 (new): A method as recited in claim 42 wherein said combination is linear.

Claim 48 (new): A method as recited in claim 42 wherein the designated variable may simultaneously comprise a plurality of values.

Claim 49 (new): A method as recited in claim 42 wherein the designated variable corresponds to a designated DNA locus.

Claim 50 (new): A method as recited in claim 42 wherein the data records are applicable to agriculture.

Claim 51 (new): A method as recited in claim 42 wherein the data records are applicable to forensic science.

Claim 52 (new): A method as recited in claim 51 where the forensic science application includes DNA analysis.

Claim 53 (new): A method as recited in claim 42 wherein the data records are applicable to space science.

Claim 54 (new): A method as recited in claim 42 wherein the data records comprise references to textual information.

Claim 55 (new): A method as recited in claim 42 wherein the value of the function is minimized.

Claim 56 (new): A method of partitioning data records in a computer into groups of approximately equal size, comprising the steps of:

(a) defining a function of a distribution of the values of a designated variable associated with the data records, wherein the function comprises a combination of measures of entropy and adjacency;

(b) partitioning the values of a designated variable into two or more groups, wherein the value of the function is determined by applying an optimization procedure; and

(c) assigning each data record to a group according to the value of the designated variable.

Claim 57 (new): A method as recited in claim 56 further comprising the step of selecting a partition from many computed solutions yielding acceptable performance.

Claim 58 (new): A method as recited in claim 56 wherein said optimization procedure results in an optimal assignment.

Claim 59 (new): A method as recited in claim 56 wherein said combination is linear.

Claim 60 (new): A method as recited in claim 56 wherein the designated variable may simultaneously comprise a plurality of values.

Claim 61 (new): A method as recited in claim 56 wherein the data records are applicable to forensic science.

Claim 62 (new): A method as recited in claim 56 wherein the designated variable corresponds to a designated DNA locus.

Claim 63 (new): A method as recited in claim 56 wherein the data records are applicable to agriculture.

Claim 64 (new): A method as recited in claim 56 wherein the data records are applicable to space science.

Claim 65 (new): A method of partitioning data for a data for a database in a computer, wherein the database is indexed using a tree of nodes, wherein the tree of nodes comprises a root node which is connected to two or more branches originating at the root node, wherein each branch terminates at a node, wherein each node other than the root node may be a non-terminal node or a leaf node, wherein each non-terminal node is connected to two or more branches originating at the non-terminal node and terminating at a node, wherein the tree-structured index comprises one or more tests associated with each non-terminal node, said method comprising the steps of:

(a) identifying naturally occurring sets of clusters in the data records of the database;

(b) defining for each identified set of clusters a test that assigns each data record within the set of clusters; and

(c) associating each test defined in step (b) with a non-terminal node and an associated set of clusters defined in step (a), and associating with each cluster within the set of clusters one branch originating at the non-terminal node, said branch forming part of one or more paths leading to leaf nodes comprising the data records assigned to the cluster by the test.

Claim 66 (new): A method as recited in claim 65 wherein said partitioning comprises partitioning of data records into groups of approximately equal size.

Claim 67 (new): A method as recited in claim 65 wherein said tests are determined by combination of entropy and adjacency.

Claim 68 (new): A method as recited in claim 67 wherein said combination is linear.

Claim 69 (new): A method as recited in claim 65 wherein the data corresponds to DNA.

Claim 70 (new): A method as recited in claim 65 wherein the database is applicable to agriculture.

Claim 71 (new): A method as recited in claim 65 wherein the database is applicable to forensic science.

Claim 72 (new): A method as recited in claim 65 wherein the database is applicable to space science.